

Xây dựng đối tượng tự động điều khiển ứng dụng trong mô phỏng huấn luyện nhảy dù

Nguyễn Trung Kiên
Học viện Kỹ thuật quân sự

TÓM TẮT

Nhảy dù là một môn thể thao mạo hiểm đồng thời cũng là nội dung huấn luyện quan trọng trong quân đội, yêu cầu người tham gia phải có bản lĩnh tâm lý vững vàng và nắm vững kỹ thuật chuyên môn. Trong thực tế, nhảy dù thường được tổ chức nhảy theo nhóm, với nhiều người nhảy cùng một đợt. Do đó, người nhảy dù cần rèn luyện khả năng phối hợp, xử lý tình huống như tránh va chạm trên không, giữ khoảng cách an toàn, hoặc xử lý khi nhiều dù cùng hạ cánh tại một khu vực. Bài báo đề xuất mô hình ứng dụng trí tuệ nhân tạo để phát triển các đối tượng nhảy dù ảo cùng tham gia quá trình nhảy dù và có khả năng tự động điều khiển dù sau khi nhảy ra khỏi máy bay để tiếp cận vị trí xác định trên mặt đất. Giải pháp này có thể được tích hợp vào các phần mềm mô phỏng huấn luyện nhảy dù, nhằm hỗ trợ huấn luyện người nhảy dù trong việc thực hành các nhiệm vụ nhảy dù theo nhóm. Kết quả nghiên cứu cho thấy việc tích hợp trí tuệ nhân tạo vào hệ thống mô phỏng huấn luyện nhảy dù là khả thi và có tiềm năng ứng dụng thực tiễn, góp phần nâng cao hiệu quả đào tạo và giảm thiểu rủi ro so với phương pháp huấn luyện truyền thống.

Từ khóa: học tăng cường, PPO, mô phỏng huấn luyện, nhảy dù, tác nhân AI

1. ĐẶT VẤN ĐỀ

Nhảy dù là một môn thể thao mạo hiểm được ưa thích trên thế giới. Trong các hoạt động quân sự, nhảy dù là phương pháp di chuyển lực lượng nhanh chóng, đặc biệt đối với lực lượng đặc công, đặc nhiệm hay các lực lượng tìm kiếm cứu nạn. Tại Việt Nam, hoạt động nhảy dù từ phương tiện bay chủ yếu được triển khai trong môi trường quân đội và được xem là một nội dung huấn luyện đặc biệt, yêu cầu mức độ chính xác và an toàn cao. Tuy nhiên, huấn luyện nhảy dù thực tế tiềm ẩn nhiều rủi ro về an toàn, đòi hỏi chi phí lớn và bị ảnh hưởng nhiều bởi điều kiện thời tiết cũng như khả năng bảo đảm phương tiện bay.

Để khắc phục các hạn chế nêu trên, nhiều quốc gia đã nghiên cứu và triển khai các hệ thống mô phỏng nhảy dù dựa trên công nghệ thực tại ảo (VR), nhằm cung cấp môi trường huấn luyện an toàn, tiết kiệm và hiệu quả hơn [1]. Các thiết bị mô phỏng này hướng đến mục tiêu hỗ trợ người học thực hành các kỹ thuật điều khiển dù trong môi trường an toàn, giảm thiểu rủi ro và tối ưu chi phí so với các phương pháp huấn luyện truyền thống. Các thiết bị mô phỏng hiện đại thường tích hợp các thành phần như phần mềm mô phỏng 3D, thiết bị tương

tác vật lý (giá treo, dây kéo, mô tơ), cảm biến và kính thực tại ảo (VR). Phần lớn các thiết bị hiện nay tập trung vào huấn luyện cá nhân, nơi học viên thực hiện toàn bộ quy trình từ lúc rời máy bay đến khi tiếp đất thông qua kính VR và các thao tác điều khiển vật lý [2 - 4]. Tại Việt Nam, công tác huấn luyện nhảy dù vẫn chủ yếu dựa vào phương pháp truyền thống: Tập trên các mô hình máy bay, hệ giá treo dù. Cho đến nay, đã có một số nghiên cứu trong nước bước đầu triển khai thiết kế, chế tạo thiết bị mô phỏng huấn luyện nhảy dù, với mục tiêu nội địa hoá thiết bị phục vụ đào tạo chuyên sâu [5]. Những thiết bị này được xây dựng trên nền tảng sử dụng kính VR, cơ cấu giá dù, hệ thống cảm biến kéo dây và phần mềm mô phỏng thời gian thực nhằm tái hiện đầy đủ thao tác điều khiển của người nhảy dù. Trong thực tế, khi tổ chức nhảy dù ngoài thực địa, các bài nhảy dù đều được tổ chức theo hình thức nhóm (nhiều người nhảy cùng lúc), đòi hỏi kỹ năng phối hợp như tránh va chạm trên không, giữ khoảng cách an toàn, hoặc xử lý khi nhiều dù cùng tiếp đất tại một khu vực. Tuy nhiên, hầu hết các thiết bị mô phỏng hiện có chưa hỗ trợ huấn luyện theo nhóm. Một số quốc gia đã đầu tư nhiều thiết

Tác giả liên hệ: Nguyễn Trung Kiên
Email: kiennt.simtech@mta.edu.vn

bị mô phỏng kết nối với nhau nhằm huấn luyện đội hình nhảy dù, nhưng đi kèm với đó là chi phí triển khai rất cao [6, 7]. Chính vì vậy, cần nghiên cứu giải pháp để thay thế đầu tư các thiết bị huấn luyện bằng đối tượng nhảy dù ảo có tính năng tương tự như người nhảy dù khi thực hành nhảy dù từ trên máy bay.

Trong những năm gần đây, trí tuệ nhân tạo (AI), đặc biệt là các phương pháp học tăng cường (Reinforcement Learning - RL) [8], đã được ứng dụng rộng rãi trong bài toán điều khiển tự động các thiết bị chuyển động trong không gian hai chiều và ba chiều (3D). Các kỹ thuật học tăng cường đã chứng minh hiệu quả vượt trội trong việc học chính sách điều khiển tối ưu thông qua quá trình tương tác lặp lại với môi trường mô phỏng và đã được áp dụng thành công trong nhiều lĩnh vực như điều hướng robot, điều khiển xe tự hành, cánh tay robot công nghiệp, cũng như phương tiện bay không người lái [9, 10]. Một điểm chung nổi bật của các hệ thống điều khiển nói trên là đặc tính phi tuyến, không xác định của môi trường vận hành, đồng thời yêu cầu điều khiển trong không gian liên tục với mức độ chính xác cao. Trong các tình huống như vậy, các thuật toán học tăng cường như DDPG, PPO, SAC và TD3 mang lại lợi thế nhờ khả năng học trực tiếp từ dữ liệu tương tác thực tế hoặc mô phỏng, giúp hệ thống tự thích nghi và cải thiện hiệu suất qua thời gian mà không cần mô hình hóa tường minh toàn bộ hệ thống vật lý [11 - 14]. Những đặc điểm này khiến học tăng cường trở thành một công cụ tiềm năng trong các hệ thống điều khiển tự động đòi hỏi tính linh hoạt, ổn định và khả năng phản ứng nhanh với thay đổi từ môi trường. Mặc dù các nghiên cứu ứng dụng học tăng cường trong điều khiển robot di động đã đạt được nhiều tiến bộ đáng kể, việc áp dụng học tăng cường cho các đối tượng có đặc tính động học phức tạp như người nhảy dù vẫn còn khó khăn. Không gian chuyển động của người nhảy dù chịu ảnh hưởng bởi trọng lực, lực cản không khí, gió và cấu trúc biến động của vòm dù, đòi hỏi mô hình học sâu không chỉ ổn định mà còn thích ứng nhanh trong không gian ba chiều. Chính vì vậy, việc nghiên cứu ứng dụng học tăng cường để xây dựng các đối tượng nhảy dù tự động trong môi trường mô phỏng ba chiều không chỉ đặt ra thách thức kỹ thuật mới, mà còn mở ra tiềm năng ứng dụng thực tiễn cao trong lĩnh vực huấn luyện nhảy dù bằng công nghệ thực tại ảo. Đặc biệt, các tác nhân điều

khiển bằng AI có thể đóng vai trò như “người nhảy dù ảo” trong huấn luyện nhóm, hỗ trợ học viên thực hành phối hợp trong các tình huống nhảy dù theo nhóm mà không cần nhiều người tham gia thực tế, từ đó giảm thiểu chi phí, tăng tính linh hoạt và đảm bảo an toàn trong huấn luyện

Trong bài báo này, tác giả đề xuất một phương pháp điều khiển người nhảy dù ảo dựa trên thuật toán học tăng cường PPO, tích hợp trong môi trường mô phỏng ba chiều thời gian thực. Khác với các nghiên cứu tập trung vào điều khiển đơn lẻ, tác nhân AI trong nghiên cứu này được thiết kế để đóng vai trò người nhảy dù trong kịch bản huấn luyện, có khả năng:

- Điều khiển quỹ đạo bay về mục tiêu;
- Tạo ra các tình huống tiếp cận nguy hiểm có kiểm soát;
- Hỗ trợ huấn luyện học viên trong các bài tập quan sát và tránh va chạm.

Trong phần tiếp theo, bài báo trình bày bài toán điều khiển dù và phương pháp sử dụng PPO trong xây dựng tác nhân AI điều khiển “người nhảy dù ảo”, đồng thời trình bày kết quả tích hợp, thử nghiệm mô hình AI với môi trường mô phỏng 3D phục vụ nội dung mô phỏng huấn luyện nhảy dù.

2. ĐỐI TƯỢNG, PHƯƠNG PHÁP VÀ NỘI DUNG NGHIÊN CỨU

2.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu bao gồm: Nội dung công tác huấn luyện nhảy dù và người nhảy dù trong quá trình thực hành nhảy dù.

2.2. Phương pháp nghiên cứu

Phương pháp nghiên cứu tài liệu huấn luyện, khảo sát thực tế công tác nhảy dù được triển khai nhằm thu thập các thông tin, dữ liệu về huấn luyện nhảy dù ở các đơn vị. Trên cơ sở đó sử dụng phương pháp phân tích, tổng hợp để xác định nội dung, yêu cầu chính của công tác huấn luyện nhảy dù. Phương pháp phân tích, thiết kế, mô hình hóa, mô phỏng được sử dụng để xây dựng các chương trình phần mềm mô phỏng nội dung và quá trình điều khiển khi nhảy dù.

2.3. Bài toán điều khiển nhảy dù

Trong huấn luyện nhảy dù, người tham gia với đầy đủ trang bị, sau khi máy bay đưa đến khu vực nhảy dù thì được lệnh nhảy ra khỏi máy bay. Sau khi rời khỏi máy bay, người nhảy dù sẽ có một giai đoạn

rori tự do. Đến độ cao yêu cầu mở dù, người nhảy dù giật vòng tay kéo để vòm dù được mở ra. Sau khi vòm dù được mở ổn định, người nhảy dù sẽ điều khiển các dây lái dù để di chuyển về vị trí xác định trên mặt đất (thường gọi là tâm bãi thả dù). Yêu cầu đặt ra: Xây dựng đối tượng “người nhảy dù ảo” có thể tự động điều khiển dù để di chuyển về vị trí tâm bãi thả dù trên mặt đất.

Bài toán điều khiển nhảy dù được xem như một bài toán điều khiển trong không gian ba chiều và có độ phức tạp cao với nhiều yếu tố vật lý tác động như trọng lực, lực cản không khí, gió, mật độ không khí ... Người nhảy dù ngồi trên giá treo dù, được gọi chung là hệ dù, điều khiển các dây lái dù. Khi không điều khiển, hệ dù sẽ tự động rơi xuống dưới tác dụng chính gây ra bởi trọng lực và lực tác động của gió. Để điều khiển dù di chuyển về vị trí đích, người nhảy dù phải tác động vào các dây lái dù để thay đổi chuyển động của hệ dù. Khi tác động vào dây lái sẽ làm biến đổi hình dạng của vòm dù, chính vì thế sẽ làm hệ dù thay đổi chuyển động. Các chuyển động chính của hệ dù bao gồm: (1) chuyển động quay (theo hướng nhìn của người nhảy dù): Quay trái, quay phải; (2) dịch chuyển: Tiến, lùi, sang trái, sang phải. Để xây dựng tác nhân AI điều khiển “người nhảy dù ảo”, thao tác điều khiển các dây lái dù cần được ánh xạ tương đương với sự thay đổi chuyển động quay và dịch chuyển.

Trong nghiên cứu này, nhóm tác giả sử dụng phương pháp học tăng cường để xây dựng tác nhân AI điều khiển người nhảy dù ảo. Học tăng cường cho phép tác nhân học chính sách điều khiển thông qua quá trình tương tác với môi trường, thay vì phải xây dựng mô hình toán học chi tiết của toàn bộ hệ thống. Trong số các thuật toán học tăng cường, thuật toán tối ưu hóa chính sách gần nhất (Proximal Policy Optimization - PPO) được lựa chọn do có ưu điểm về tính ổn định, dễ huấn luyện và phù hợp với bài toán điều khiển liên tục. Thuật toán PPO sử dụng cơ chế clipping để giới hạn mức độ cập nhật chính sách ở mỗi bước, điều này đặc biệt quan trọng với bài toán điều khiển nhảy dù có không gian trạng thái liên tục và đòi hỏi tính ổn định cao. So với các thuật toán khác như DDPG hoặc TD3 yêu cầu điều chỉnh nhiều siêu tham số và dễ gặp vấn đề mất ổn định, PPO mang lại độ ổn định vượt trội nhờ cơ chế kiểm soát cập nhật chính sách.

Trong quá trình huấn luyện, tác nhân có khả năng học cách thức điều khiển các dây dù (chính sách

điều khiển - policy) thông qua việc tương tác với môi trường và tối ưu hóa hàm mục tiêu dựa trên phần thưởng (reward) nhận được. Phần thưởng này được thiết kế nhằm khuyến khích hành động điều khiển tiến gần điểm đích hoặc duy trì hướng nhìn về phía đích đến. Bài báo triển khai thuật toán PPO để huấn luyện tác nhân AI, với mục tiêu tối đa hóa phần thưởng tích lũy trong suốt quá trình điều khiển, từ đó nâng cao hiệu quả tiếp cận điểm đích của đối tượng tự động điều khiển dù.

- Môi trường mô phỏng: Được xây dựng dựa trên mô hình động lực học dù [5], tích hợp các yếu tố vật lý như trọng lực, lực cản không khí và tác động của gió.
- Thu thập kinh nghiệm: Tác nhân thực hiện hành động ngẫu nhiên ban đầu, quan sát kết quả (trạng thái mới, phần thưởng) và lưu trữ các bộ dữ liệu.

Toàn bộ dữ liệu huấn luyện được sinh tự động từ môi trường mô phỏng, không yêu cầu dữ liệu nhảy dù thực tế từ con người.

2.4. Môi trường hệ dù

Môi trường hệ dù được đặc tả như sau:

- Không gian tính toán: \mathbb{R}^3
- Gia tốc trọng trường: $g = 9.81 \text{ m/s}^2$
- Khối lượng người nhảy dù: M
- Khối lượng dù: m
- Diện tích vòm dù: S
- Vận tốc máy bay: (v_{mx}, v_{my}, v_{mz})
- Vị trí người nhảy dù ban đầu: (x_0, y_0, z_0)
- Góc hướng nhìn ban đầu: ψ_0
- Vị trí điểm tâm bãi: $\vec{G}(x_g, y_g, z_g)$
- Vận tốc gió: (w_{0x}, w_{0y}, w_{0z})

Trong đó, mô hình được xây dựng dựa trên động lực học của dù tròn D6, với các phương trình chuyển động tính đến trọng lực, lực cản không khí và tác động của gió. Mô hình giả định vòm dù đã mở hoàn toàn và ổn định, không xét đến giai đoạn trung gian khi mở dù và các hiện tượng dao động của vòm dù, cũng như chưa tính đến các hiện tượng vật lý phức tạp như xoáy khí động, hoặc tương tác phi tuyến giữa người và dù.

2.5. Trạng thái hệ dù

Trạng thái hiện tại của dù được thể hiện qua các thông số sau:

- Vị trí của hệ dù: $\vec{P}_t(x, y, z)$
- Vận tốc dịch chuyển của hệ dù: (v_{dx}, v_{dy}, v_{dz})
- Góc quay của hệ dù: (φ, θ, ψ)
- Vận tốc góc quay của hệ dù: (p, q, r)

- Tốc độ gió hiện tại: (w_n, w_e, w_d)

Như vậy, mỗi trạng thái của dù được biểu diễn dưới dạng vector $s_t \in \mathbb{R}^{15}$.

2.5.1. Hành động điều khiển dù

Đối với loại dù D6 có vòm dù dạng hình nửa cầu, đối xứng và có 6 dây lái dù, bao gồm hai dây lái hướng và 4 dây treo người. Việc kéo các dây điều khiển sẽ làm thay đổi hình dạng cục bộ của vòm dù, từ đó tạo ra sự chênh lệch áp suất khí động giữa các vùng, dẫn đến sự thay đổi lực nâng, lực cản và mô-men quay từ đó làm cho hệ dù dịch chuyển. Ví dụ, nếu kéo dây phía trước bên trái, vòm dù sẽ nghiêng về phía trước, bên trái và tạo ra mô-men quay quanh trục đứng, làm thay đổi hướng bay, tiến về phía trước, bên trái. Nếu kéo đồng thời cả hai dây treo người phía bên trái, hệ dù sẽ dịch chuyển sang bên trái. Tác động tương tự đối với hệ dù nếu người nhảy dù kéo các dây còn lại.

Hai dây lái hướng có tác dụng kiểm soát cửa gió của vòm dù. Khi người nhảy kéo dây lái hướng, cửa gió sẽ đóng hoặc mở, làm thay đổi sự phân phối không khí trong vòm dù. Việc thay đổi sự phân phối không khí này tạo ra momen khí động tác dụng lên vòm dù, làm quay vòm dù, qua đó làm thay đổi hướng nhìn của hệ dù. Khi kéo dây lái bên trái, hệ dù sẽ quay sang trái và ngược lại. Tóm lại, các hành động tác động lên dây lái dù sẽ dẫn đến thay đổi trạng thái của dù: Quay trái/phải, tiến/lùi, hoặc dịch chuyển trái/phải.

Nhằm giảm thiểu độ phức tạp của không gian hành động và tăng tính khả thi trong huấn luyện, nghiên cứu tiến hành đơn giản hóa không gian hành động điều khiển 6 dây lái dù thành tập không gian hành động 3 bậc tự do để điều khiển dù: $A = \{a_i | a_i \in [-1, 1], i = 0, 1, 2\} \in [-1, 1]^3$, trong đó, a_0 đại diện cho tốc độ quay trái/phải; a_1 đại diện cho dịch chuyển tiến/lùi; a_2 đại diện cho dịch chuyển trái/phải.

Việc giảm từ 6 dây điều khiển độc lập xuống 3 bậc tự do được thực hiện dựa trên phân tích động học của dù tròn D6. Trong thực tế, các động tác điều khiển dù cơ bản của người nhảy dù có thể được tổng hợp thành 3 nhóm chuyển động chính: (1) quay trái/phải; (2) tiến/lùi; (3) dịch chuyển ngang trái/phải.

Ánh xạ từ không gian hành động 3D sang điều khiển 6 dây được thực hiện như sau:

- Quay trái/phải (a_0): Điều khiển 2 dây lái hướng theo nguyên tắc đối xứng.

- Tiến/lùi (a_1): Điều khiển cặp dây treo người phía trước/sau.

- Dịch ngang (a_2): Điều khiển cặp dây treo người trái/phải.

Việc đơn giản hóa này giúp: (1) giảm 50% số chiều của không gian hành động, từ đó tăng tốc độ hội tụ của thuật toán; (2) phù hợp với cách điều khiển tự nhiên của người nhảy dù trong thực tế; (3) vẫn đảm bảo khả năng thực hiện đầy đủ các động tác điều khiển cần thiết.

2.5.2. Phần thưởng cho hành động điều khiển dù

Mục tiêu của quá trình điều khiển dù tự động là tiếp cận đến vị trí điểm đích. Vì vậy, phần thưởng cho hành động điều khiển dù được thể hiện qua khoảng cách và hướng nhìn về vị trí điểm đích.

Thông thường, độ cao máy bay bay nhảy dù trung bình là 800 m, vận tốc rơi trung bình của dù là 5 m/s, thời gian rơi trung bình khi đó là 160 s. Sau khi mở dù, người nhảy dù phải nhanh chóng xác định vị trí điểm đích và điều khiển dù để quay về hướng điểm đích. Tốc độ quay trung bình là 180° trong vòng 60 s. Do đó, yêu cầu người nhảy dù phải quay được về hướng điểm đích trong vòng 60 s.

Hàm phần thưởng $R(s_t, a_t)$ được thiết kế để hướng dẫn tác nhân học cách điều khiển quay để hướng hệ dù về phía điểm đích trong giai đoạn đầu và sau đó điều khiển dịch chuyển để tiếp cận điểm đích trong không gian 3D. Phần thưởng bao gồm hai thành phần: Thưởng theo hướng và thưởng theo khoảng cách.

$$R(s_t, a_t) = \rho r_{heading} + \varphi r_{dist}, \rho, \varphi \in \mathbb{R} \quad (1)$$

Trong đó, ρ, φ là hai trọng số, thể hiện mức độ ưu tiên của từng mục tiêu trong quá trình huấn luyện. Việc điều chỉnh các hệ số này cho phép kiểm soát hướng tập trung của chính sách điều khiển. Thời gian đầu (trong vòng 60s), ρ nhận giá trị lớn để mô hình ưu tiên điều khiển về hướng di chuyển. Sau đó, φ nhận giá trị lớn để mô hình tập trung vào việc điều khiển tiếp cận mục tiêu. Các cặp giá trị (ρ, φ) được chọn là:

$$(\rho, \varphi) = \begin{cases} (10.0, 0.1) & \text{nếu } t \leq 60 \\ (0.1, 10.0) & \text{ngược lại} \end{cases} \quad (2)$$

Các trọng số này điều khiển sự cân bằng giữa hai mục tiêu: (1) điều chỉnh hướng nhìn về đích và (2) tiếp cận đích. Trong giai đoạn đầu (60 giây đầu), $\rho = 10.0$ được ưu tiên để tác nhân tập trung học cách quay về hướng đích nhanh chóng. Sau đó,

$\varphi = 10.0$ được áp dụng để tập trung vào kỹ năng tiếp cận đích. Việc thiết lập này dựa trên thực tế huấn luyện nhảy dù: Người nhảy dù thường phải định hướng trước khi điều khiển tiến về đích. Giá trị $\rho = 10$ được chọn sau khi thử nghiệm với các giá trị từ 5 đến 20. Kết quả cho thấy $\rho = 10$ tạo ra gradient phần thưởng đủ mạnh để hướng dẫn tác nhân nhưng không quá nhạy gây dao động trong quá trình học. Với ρ nhỏ hơn ($\rho < 5$), tác nhân học chậm hơn trong việc điều chỉnh hướng. Với ρ lớn hơn ($\rho > 15$), xuất hiện hiện tượng dao động do phần thưởng thay đổi quá mạnh giữa các bước. Tương tự đối với hệ số φ .

+ Phần thưởng theo hướng

Tác nhân cần điều khiển dù để quay mặt hướng về phía điểm đích (xét trong mặt phẳng ngang xOy). Vec-tơ hướng nhìn: $\vec{v}_{face} = [\cos\psi, \sin\psi]$. (3)

Vec-tơ hướng tới điểm đích: $\vec{v}_{goal} = \vec{G} - \vec{P}_t$. (4)

Giá trị độ lệch góc giữa hướng nhìn và hướng điểm đích: $\gamma = \arccos\left(\frac{\vec{v}_{face} \cdot \vec{v}_{goal}}{\|\vec{v}_{face}\| \|\vec{v}_{goal}\|}\right)$. (5)

Phần thưởng theo hướng được tính theo công thức sau:

$$r_{heading} = \begin{cases} +10 & \text{nếu } |\gamma| \leq 10^0 \\ +\lambda \cos\gamma & \text{nếu } 10^0 < |\gamma| \leq 60^0 \\ -5 & \text{nếu } |\gamma| > 60^0 \end{cases} \quad (6)$$

trong đó: λ là hệ số thưởng, chọn $\lambda = 2.0$.

+ Phần thưởng theo khoảng cách

Khoảng cách 3D từ vị trí hiện tại của hệ dù P_t đến điểm đích G : $d_t = \|\vec{P}_t - \vec{G}\|$. Hàm phần thưởng theo khoảng cách r_{dist} được tính theo công thức sau:

$$r_{dist} = \begin{cases} -\alpha \cdot d_t & , \text{ khi chưa tiếp đất} \\ +50 & , \text{ tiếp đất nếu } d_t \leq 20 \text{ m} \\ -10 & , \text{ tiếp đất nếu } d_t > 20 \text{ m} \end{cases} \quad (7)$$

Giá trị α cần được chọn để thiết kế hàm phần thưởng có cấu trúc định hình tốt hơn cho tác nhân bằng cách thêm một lượng phạt nhỏ ở mỗi bước tùy theo khoảng cách đến đích, giúp tác nhân học nhanh hơn, ổn định hơn. Khi tiếp đất đúng khu vực điểm đích thì sẽ được thưởng lớn, ngoài ra thì bị phạt. Khi chưa tiếp đất, tổng phần thưởng định hình trong quá trình huấn luyện là:

$$R_{shaping} = -\alpha \sum_{t=0}^{n-1} d_t.$$

Trong quá trình học tăng cường, phần thưởng định hình $R_{shaping}$ đóng vai trò định hướng tác nhân tiến dần đến mục tiêu. Tuy nhiên, nếu phần

thưởng này không được kiểm soát về biên độ và độ phân tán, có thể gây mất ổn định khi cập nhật chính sách điều khiển. Do đó, cần kiểm soát độ lệch chuẩn của phần thưởng định hình để giữ ổn định cho đạo hàm lan truyền ngược (gradient) trong quá trình cập nhật chính sách. Qua thực nghiệm, giá trị hệ số được lựa chọn: $\alpha = 0.005$, đây là giá trị đủ nhỏ để không ảnh hưởng tới phần thưởng chính nhưng đủ lớn để tạo ra gradient hướng về điểm đích.

2.5.3. Chính sách điều khiển dù

Trong khuôn khổ học tăng cường, chính sách là một hàm ánh xạ từ trạng thái của môi trường sang hành động mà tác nhân thực hiện nhằm tối ưu hóa phần thưởng tích lũy theo thời gian. Trong bài toán điều khiển nhảy dù, chính sách điều khiển đóng vai trò trung tâm trong việc quyết định cách tác nhân điều chỉnh các dây dù để định hướng và tiếp cận điểm đích một cách hiệu quả.

Chính sách điều khiển dù được biểu diễn dưới dạng một hàm xác suất tham số hóa $\pi_{\theta}(a|s)$, với s là trạng thái của hệ dù tại một thời điểm ($s \in \mathbb{R}^{15}$) và a là hành động mà tác nhân lựa chọn ($a \in \mathbb{R}^3$), còn θ là các tham số của mạng chính sách. Dựa vào đó, chính sách học được sẽ quyết định giá trị hành động điều khiển dù tương ứng, bao gồm: Quay trái/phải, tiến/lùi, hoặc dịch chuyển trái/phải.

Chính sách này được huấn luyện theo phương pháp tối ưu hóa chính sách gần nhất PPO. Mô hình chính sách được sử dụng trong quá trình huấn luyện là mạng nơ-ron sâu với kiến trúc ba lớp ẩn tương ứng với số nút mạng nơ-ron trong từng lớp là 256, 128 và 64. Quá trình huấn luyện chính sách được tiến hành theo các vòng lặp: Tác nhân tương tác với môi trường để thu thập dữ liệu trạng thái, sau đó cập nhật chính sách thông qua việc tối đa hóa phần thưởng kỳ vọng. Phần thưởng đã được thiết kế để khuyến khích tác nhân hướng về phía điểm đích và rút ngắn khoảng cách đến điểm đích tại mỗi bước điều khiển. Cụ thể, các hành động có xu hướng cải thiện hướng nhìn hoặc làm giảm khoảng cách về điểm đích sẽ nhận được phần thưởng dương, trong khi các hành động lệch hướng hoặc tăng khoảng cách sẽ bị phạt. Kết quả của quá trình huấn luyện chính sách là một mô hình điều khiển dù có khả năng phản ứng linh hoạt với môi trường không xác định, từ đó có thể tiếp cận đến điểm đích.

2.5.4. Lưu mô hình tốt nhất trong quá trình huấn luyện

Trong quá trình huấn luyện tác nhân điều khiển dù bằng PPO, việc theo dõi và lưu lại chính sách tốt nhất đóng vai trò quan trọng nhằm đảm bảo chất lượng mô hình. Việc lưu mô hình tốt nhất không chỉ phục vụ cho đánh giá chất lượng mà còn là cơ sở để triển khai áp dụng trong các hệ thống mô phỏng huấn luyện nhảy dù thực tế.

Một trong những thách thức lớn trong huấn luyện tác nhân tự động điều khiển dù là đảm bảo chính sách học được có thể tổng quát hóa tốt sang những tình huống chưa có trong quá trình huấn luyện. Để làm được như vậy, sau mỗi vòng huấn luyện, các thông số trạng thái được khởi tạo ngẫu nhiên, trong đó quan trọng nhất là các thông số. Các giá trị được thiết lập để khởi tạo ngẫu nhiên: Vị trí nhảy $\{(x_0, y_0, z_0) | (x_0, y_0) \in [-300, 300]^2, z_0 \in [500, 1000]\}$, và góc hướng nhìn ban đầu $\{\psi_0 | \psi_0 \in [0, 2\pi]\}$.

2.5.5. Tích hợp tác nhân AI trong kịch bản huấn luyện nhóm

Khác với các nghiên cứu tập trung vào điều khiển đơn lẻ, trong hệ thống đề xuất, tác nhân AI được tích hợp như một thành phần tham gia trực tiếp vào kịch bản huấn luyện. Tác nhân AI đóng vai trò là người nhảy dù ảo, có khả năng tạo ra các tình huống tiếp cận gần hoặc giao thoa quỹ đạo với học viên.

Trong các kịch bản huấn luyện nhóm, học viên quan sát quỹ đạo chuyển động của tác nhân AI, nhận biết nguy cơ va chạm và thực hành điều chỉnh dù để tránh va chạm trong không gian ba chiều. Các tình huống này được thiết kế có kiểm soát nhằm đảm bảo an toàn, đồng thời nâng cao kỹ năng phản xạ và nhận thức không gian cho học viên. Việc tích hợp tác nhân AI trong kịch bản huấn luyện cho phép hệ thống mô phỏng tái hiện các tình huống huấn luyện khó tổ chức trong thực tế, đồng thời giảm chi phí và rủi ro so với huấn luyện

ngoài thao trường.

Trong hệ thống huấn luyện đề xuất, tác nhân AI không chỉ thực hiện nhiệm vụ điều khiển tự động mà còn đóng vai trò là một người nhảy dù ảo tham gia trực tiếp vào kịch bản huấn luyện. Tác nhân này được thiết kế để mô phỏng hành vi của con người trong điều kiện thực tế, từ đó tạo ra các tình huống huấn luyện phong phú và có tính thực tiễn cao.

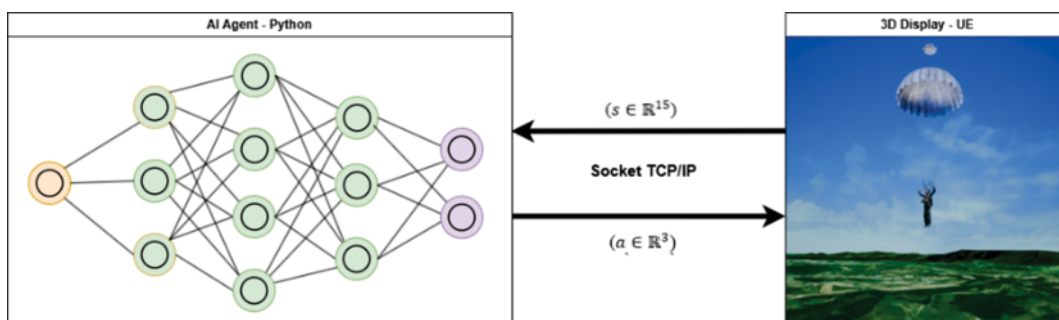
Thông qua việc quan sát chính sách điều khiển và phản ứng của tác nhân AI, học viên có thể rèn luyện kỹ năng nhận biết quỹ đạo, đánh giá nguy cơ và thực hành tránh va chạm trong không gian ba chiều một cách an toàn và hiệu quả. Điều này cho phép mô phỏng các tình huống huấn luyện nhóm, trong đó học viên có thể quan sát, tương tác và thực hành tránh va chạm với các đối tượng nhảy dù ảo, góp phần nâng cao hiệu quả và tính an toàn của công tác huấn luyện.

3. KẾT QUẢ NGHIÊN CỨU VÀ BÀN LUẬN

3.1. Môi trường mô phỏng và cấu hình thử nghiệm

Các thử nghiệm được thực hiện trong môi trường mô phỏng ba chiều thời gian thực, được xây dựng trên nền tảng phần mềm mô phỏng 3D. Môi trường này cho phép mô phỏng toàn bộ quá trình nhảy dù từ thời điểm mở dù đến khi tiếp đất, bao gồm các yếu tố chuyển động trong không gian ba chiều, ảnh hưởng của gió và địa hình khu vực hạ cánh.

Mô hình hệ thống tự động điều khiển nhảy dù được thiết kế theo kiến trúc ba thành phần chính: Phần mềm AI tự động tính toán điều khiển chạy trên nền Python (AI-Python), phần mềm mô phỏng 3D triển khai trên nền tảng thư viện 3D (3D Display-UE) và modul kết nối mạng TCP/IP. Mục tiêu của cấu trúc tích hợp này là đảm bảo việc điều khiển tự động của tác nhân AI có thể được kiểm tra, đánh giá và trực quan hóa trong một môi trường mô phỏng thực tế, có độ phức tạp vật lý cao. Mô hình kiến trúc tích hợp được mô tả trong Hình 1.



Hình 1. Mô hình kiến trúc hệ thống tích hợp

Tác nhân AI được tích hợp trực tiếp vào môi trường mô phỏng thông qua cơ chế trao đổi dữ liệu thời gian thực. Tại mỗi bước mô phỏng, môi trường cung cấp thông tin trạng thái cho tác nhân AI, tác nhân tính toán hành động điều khiển và gửi lệnh điều khiển trở lại môi trường để cập nhật trạng thái tiếp theo. Cơ chế này cho phép đánh giá chính xác khả năng điều khiển của tác nhân AI trong điều kiện gần với huấn luyện thực tế.

Các kịch bản thử nghiệm được thiết kế với nhiều điều kiện khởi tạo khác nhau về vị trí, độ cao và hướng rơi ban đầu, nhằm đánh giá khả năng thích nghi của mô hình điều khiển trong các tình huống đa dạng.

3.2. Quy trình huấn luyện và đánh giá tác nhân AI

Quá trình huấn luyện tác nhân AI được thực hiện thông qua nhiều vòng mô phỏng liên tiếp. Trong mỗi vòng, tác nhân AI thực hiện nhiệm vụ điều khiển người nhảy dù tiếp cận khu vực hạ cánh mục tiêu dựa trên chính sách điều khiển học được. Thông tin phản hồi từ môi trường, bao gồm giá trị hàm thưởng và trạng thái tiếp theo, được sử dụng để cập nhật chính sách điều khiển.

Sau khi hoàn thành quá trình huấn luyện, tác nhân AI được đưa vào các kịch bản thử nghiệm độc lập để đánh giá khả năng điều khiển. Việc đánh giá tập trung vào các tiêu chí chính như độ chính xác khi tiếp cận điểm hạ cánh, độ ổn định của quỹ đạo bay

và khả năng duy trì hành vi điều khiển hợp lý trong suốt quá trình hạ cánh.

Bên cạnh các kịch bản điều khiển đơn lẻ, tác nhân AI còn được sử dụng trong các kịch bản huấn luyện nhóm, trong đó AI đóng vai người nhảy dù ảo tương tác với học viên huấn luyện.

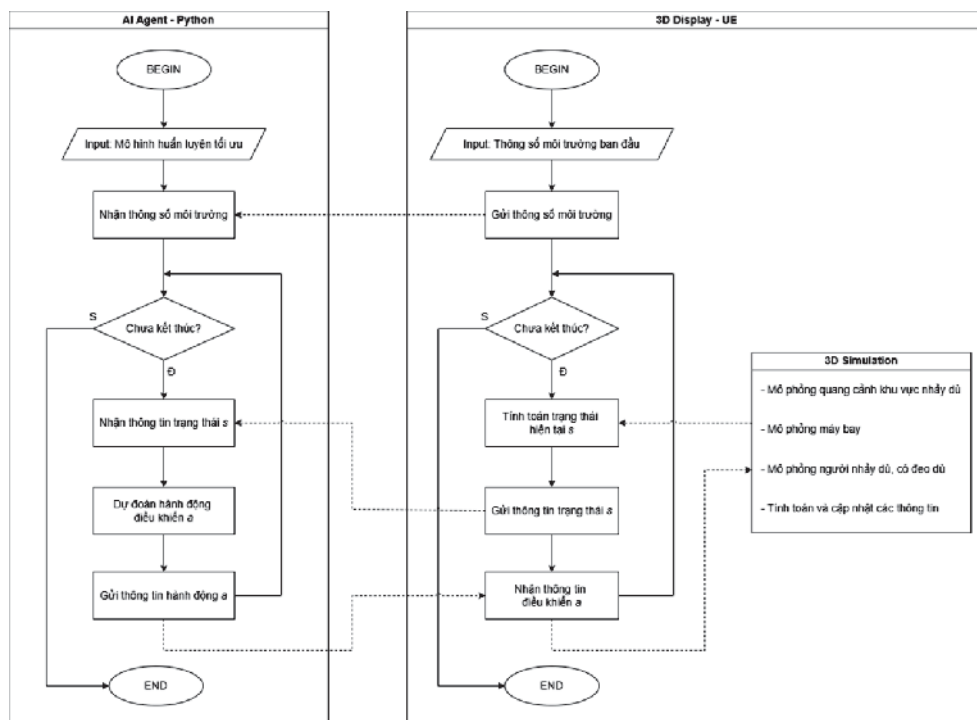
3.3. Mô hình hoạt động của hệ thống

Phần mềm AI sử dụng mô hình PPO đã được huấn luyện tốt nhất, nhận thông tin trạng thái hiện tại của hệ dù và sử dụng để dự đoán hành động điều khiển dù tối ưu. Mỗi vòng lặp tương tác gồm ba bước:

- Nhận trạng thái của hệ dù tại thời điểm hiện tại ($s \in \mathbb{R}^{15}$) từ phần mềm mô phỏng 3D.
- Dự đoán hành động ($a \in \mathbb{R}^3$) thông qua mạng chính sách PPO đã huấn luyện.
- Trả về kết quả hành động điều khiển ($a \in \mathbb{R}^3$) cho phần mềm mô phỏng 3D để cập nhật, tính toán.

Phần mềm mô phỏng 3D đảm nhận hai tiến trình chính: Tiến trình trao đổi thông tin với phần mềm AI và tiến trình mô phỏng các đối tượng môi trường 3D (quang cảnh, máy bay, người, dù...) với các thông tin cập nhật từ AI. Tiến trình trao đổi thông tin với phần mềm AI cập nhật trung bình 30 Hz. Tiến trình mô phỏng 3D có thời gian cập nhật rất nhanh (60 - 90 Hz).

Sơ đồ khối các nội dung thực hiện chính của hai phần mềm được mô tả trong Hình 2.



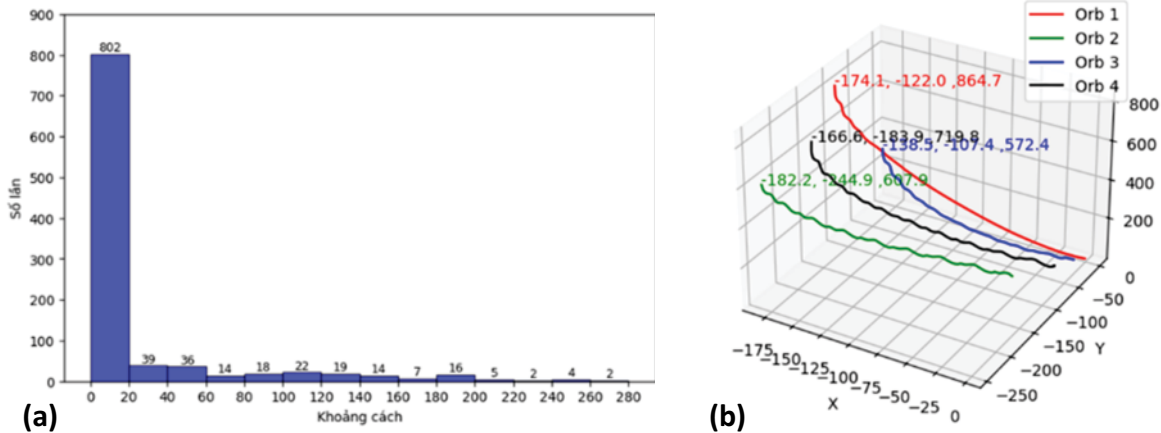
Hình 2. Sơ đồ khối các nội dung thực hiện chính của hệ thống

3.4. Kết quả thử nghiệm

Các thử nghiệm được tiến hành trên máy tính với cấu hình bộ vi xử lý Intel Core i7-13620H, bộ nhớ RAM 32GB, ổ cứng SSD dung lượng 1TB, card đồ họa GTX 1650. Với cấu hình này, quá trình huấn

luyện mô hình học tăng cường diễn ra ổn định với tốc độ nhanh.

Kết quả thử nghiệm mô hình huấn luyện tự động điều khiển:

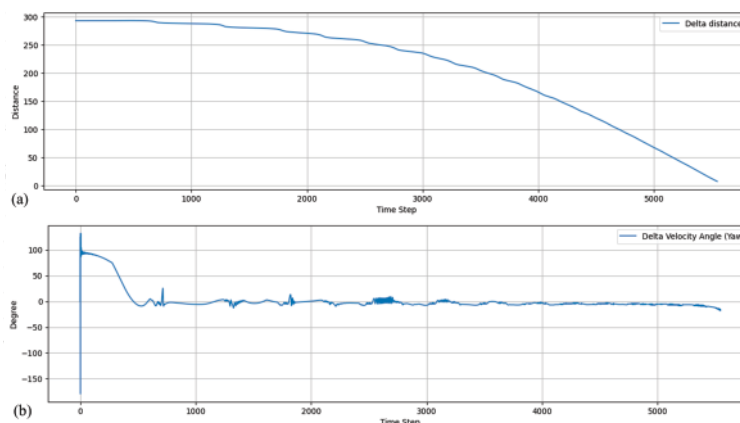


Hình 3. Kết quả thử nghiệm của mô hình sau khi huấn luyện: (a) Biểu đồ độ lệch khoảng cách so với điểm đích của các lần nhảy, (b) Quỹ đạo chuyển động của các lần nhảy dù

Hình 3 trình bày thống kê độ lệch của 1,000 lần nhảy được điều khiển bởi tác nhân AI, với các giá trị ngẫu nhiên về vị trí và góc nhảy ban đầu. Kết quả cho thấy hơn 80% lần tiếp đất nằm trong khoảng cách dưới 20 m so với điểm đích, điều này chứng tỏ tác nhân đã học được một chính sách điều khiển hiệu quả. Các lần tiếp đất còn lại phân bố thưa dần ở các khoảng cách xa hơn, với khoảng 5% số lần nhảy có độ lệch vượt quá 150 m, chứng tỏ tỷ lệ lỗi lớn là rất thấp. Cụ thể: Khoảng 56% (28 trường hợp) bị ảnh hưởng bởi điều kiện biên, gió mạnh; 44% (22 trường hợp) hạn chế mạng chính sách do góc lệch lớn từ khoảng 75° - 105°.

lần nhảy từ các vị trí khởi đầu khác nhau. Kết quả cho thấy các tác nhân đều có xu hướng điều khiển quỹ đạo tiến dần về vị trí điểm đích tại tọa độ (0, 0, 0). Trong số đó, các quỹ đạo Orb 1, Orb 3 và Orb 4 thể hiện khả năng định hướng tiếp đất tốt khi điểm hạ cánh nằm gần khu vực điểm đích. Đặc biệt, quỹ đạo Orb 3 bắt đầu từ độ cao thấp hơn so với các trường hợp khác nhưng vẫn đạt hiệu quả cao do khoảng cách ngang tới đích là tương đối ngắn. Ngược lại, quỹ đạo Orb 2 có vị trí nhảy ban đầu xa hơn và độ cao thấp, dẫn đến vị trí tiếp đất lệch xa nhiều so với điểm đích; tuy nhiên, hướng chuyển động vẫn cho thấy xu hướng hội tụ về điểm đích. Kết quả này cho thấy tính ổn định và khả năng tổng quát hóa của chính sách điều khiển học được trong các tình huống thay đổi giá trị khởi tạo ban đầu ngẫu nhiên của các thông số môi trường.

Hình 4 thể hiện các quỹ đạo chuyển động trong không gian ba chiều của tác nhân AI được điều khiển bởi mô hình học tăng cường PPO sau khi huấn luyện. Mỗi đồ thị quỹ đạo tương ứng với một



Hình 4. (a) Sự thay đổi về khoảng cách theo mặt phẳng nằm ngang, (b) Thay đổi về độ lệch giữa hướng của hệ dù và hướng về điểm đích trong quá trình điều khiển

3.5. Kết quả thử nghiệm phần mềm mô phỏng 3D

Phần mềm mô phỏng 3D chịu trách nhiệm hiển thị các nội dung trong không gian ảo ba chiều. Hình 5 mô tả một số hình ảnh mô phỏng các đối tượng trong không gian 3D, gồm có địa hình, sân bay, máy bay ...

Máy bay sẽ cất cánh từ sân bay, chuyển động theo tuyến cố định (Hình 5a). Trong quá trình đó, người nhảy dù đứng tại cửa máy bay, chuẩn bị sẵn sàng nhảy (Hình 5b). Khi máy bay đến một vị trí xác định, người nhảy dù nhảy ra khỏi cửa máy bay (Hình 5c).



Hình 5. Mô phỏng các đối tượng trong không gian 3D. (a) Máy bay trên sân bay. (b) Người nhảy dù đứng tại cửa máy bay, chuẩn bị nhảy. (c) Người nhảy dù nhảy ra khỏi máy bay

Khi hai phần mềm kết nối với nhau, các thông tin về môi trường được gửi từ phần mềm mô phỏng 3D sang phần mềm AI như vị trí điểm đích, vận tốc gió ... Khi người nhảy dù mở dù, thông tin của hệ dù như vị trí, vận tốc dịch chuyển, góc quay ... được gửi sang phần mềm AI. Sau khi nhận các

thông tin điều khiển từ phần mềm AI truyền sang, phần mềm mô phỏng 3D tính toán và cập nhật trạng thái của hệ dù (người và dù): Hướng nhìn, hướng di chuyển, tốc độ di chuyển ... Hình 6 mô tả quá trình “người nhảy dù ảo” di chuyển về phía điểm đích.



Hình 6. Quá trình “người nhảy dù ảo” di chuyển về phía điểm đích (điểm dấu T)

3.6. Kịch bản huấn luyện nhóm với tác nhân AI

Để làm rõ vai trò của tác nhân AI trong huấn luyện, các kịch bản huấn luyện nhóm được xây dựng, trong đó học viên và tác nhân AI cùng tham gia trong một môi trường mô phỏng. Tác nhân AI được điều khiển nhằm tạo ra các tình huống tiếp cận gần hoặc giao thoa quỹ đạo với học viên, nhưng vẫn đảm bảo khoảng cách an toàn.

Trong các kịch bản này, học viên có nhiệm vụ quan sát quỹ đạo chuyển động của tác nhân AI, nhận biết nguy cơ va chạm và thực hành điều chỉnh dù để tránh va chạm. Việc sử dụng tác nhân AI cho phép tái hiện các tình huống huấn luyện phức tạp mà khó có thể tổ chức an toàn trong huấn luyện

thực tế ngoài thao trường.

3.7. Thảo luận, đánh giá

Việc áp dụng học tăng cường RL trong bài toán điều khiển tự động nhảy dù, kết hợp với môi trường mô phỏng 3D thời gian thực, đã mang lại những tiềm năng đáng kể trong việc tự động hóa quá trình điều khiển và nâng cao tính linh hoạt của hệ thống trước các điều kiện môi trường thay đổi. Sự kết hợp này không chỉ tạo ra một hệ thống có khả năng điều khiển chính xác và ổn định trong các tình huống thực tế mà còn mở ra cơ hội ứng dụng trong các hệ thống huấn luyện mô phỏng hiện đại. Kết quả huấn luyện với thuật toán PPO

cho thấy tác nhân AI có thể học được chính sách điều khiển hiệu quả thông qua quá trình tương tác với môi trường mô phỏng. Mặc dù mô hình phải xử lý với các điều kiện ban đầu ngẫu nhiên (như vị trí, độ cao và góc nhìn), nhưng tác nhân vẫn duy trì khả năng điều hướng chính xác về mục tiêu. Điều này khẳng định khả năng của phương pháp học tăng cường trong việc tìm ra các chính sách điều khiển phi tuyến, thường rất khó mô hình hoá đối với các thuật toán truyền thống. Hơn nữa, kết quả thử nghiệm thực tế trong mô phỏng cho thấy tỷ lệ thành công của các lần tiếp đất gần điểm đích đạt mức cao và ổn định. Điều này chứng tỏ rằng hành vi điều khiển của tác nhân vẫn duy trì sự ổn định trong nhiều điều kiện môi trường thay đổi, từ đó nâng cao tính khả thi của hệ thống trong thực tế.

So sánh với phương pháp điều khiển rule-based đơn giản: Sử dụng các quy tắc điều khiển đơn giản (luôn quay về hướng đích với tốc độ cố định, sau đó tiến thẳng). Kết quả cho thấy tỷ lệ tiếp đất trong bán kính < 50 m chỉ đạt 67.4%, thấp hơn PPO (93.7%) đến 26.3%. Điều này chứng tỏ PPO học được chính sách điều khiển phức tạp hơn, có khả năng thích nghi với điều kiện gió và tối ưu hóa quỹ đạo. Với phương pháp DDPG, huấn luyện tác nhân với thuật toán DDPG trong cùng điều kiện. Kết quả cho thấy DDPG đạt tỷ lệ thành công 88.2% (trong bán kính < 50 m). Hơn nữa, quá trình huấn luyện DDPG kém ổn định hơn với độ dao động lớn hơn trong phần thưởng.

Một trong những điểm mạnh của hệ thống là việc kết hợp mô hình AI với nền tảng mô phỏng 3D. Mô phỏng 3D không chỉ cho phép tạo ra các kịch bản phức tạp với các thông số vật lý linh hoạt, như gió, trọng lực và địa hình, mà còn giúp kiểm tra tính hiệu quả của mô hình trong các điều kiện khác nhau. Việc tách biệt giữa quá trình huấn luyện và mô phỏng cũng tạo ra một lợi thế đáng kể, khi mô hình AI có thể được huấn luyện độc lập trong môi trường Python và sau đó được tích hợp vào hệ thống mô phỏng đồ họa. Điều này không chỉ giúp dễ dàng trong việc hiệu chỉnh và nâng cấp mô hình mà còn giảm thiểu sự gián đoạn trong quá trình huấn luyện.

Mặc dù kết quả thử nghiệm cho thấy giải pháp này mang lại hiệu quả khả quan, nhưng vẫn tồn tại một số hạn chế. Quá trình huấn luyện PPO yêu cầu một số lượng tương tác lớn, điều này làm gia

tăng chi phí tính toán và thời gian huấn luyện. Bên cạnh đó, việc duy trì kết nối liên tục giữa mô hình AI và môi trường mô phỏng có thể gặp vấn đề về độ trễ hoặc mất đồng bộ, đặc biệt trong các môi trường yêu cầu tính toán thời gian thực. Điều này làm tăng độ phức tạp và có thể ảnh hưởng đến độ tin cậy của hệ thống khi triển khai vào huấn luyện thực tế. Để giải quyết các vấn đề này, việc tối ưu hóa tài nguyên hệ thống, cải thiện kỹ thuật truyền dữ liệu và đồng bộ giữa các thành phần sẽ là yếu tố quan trọng để nâng cao hiệu quả và độ ổn định của hệ thống trong các ứng dụng thực tế.

4. KẾT LUẬN

Bài báo đã trình bày mô hình tự động điều khiển dựa trên học tăng cường và ứng dụng trong môi trường mô phỏng huấn luyện nhảy dù ba chiều. Trên cơ sở tích hợp tác nhân trí tuệ nhân tạo với môi trường mô phỏng thời gian thực, nghiên cứu đã xây dựng được mô hình điều khiển có khả năng điều chỉnh quỹ đạo bay của người nhảy dù, đảm bảo tiếp cận khu vực tiếp đất một cách ổn định và phù hợp với yêu cầu huấn luyện.

Kết quả thử nghiệm mô phỏng cho thấy tác nhân AI có khả năng học chính sách điều khiển hiệu quả thông qua quá trình tương tác với môi trường, không yêu cầu mô hình hóa chi tiết toàn bộ động lực học của hệ thống. Đặc biệt, việc sử dụng tác nhân AI như một người nhảy dù ảo tham gia trực tiếp vào các kịch bản huấn luyện đã góp phần mở rộng phạm vi ứng dụng của mô hình, từ điều khiển tự động đơn lẻ sang hỗ trợ huấn luyện nhóm trong điều kiện an toàn và có kiểm soát. Việc tích hợp trí tuệ nhân tạo vào môi trường huấn luyện cho phép tạo ra các tình huống huấn luyện linh hoạt, có thể lặp lại nhiều lần và giảm thiểu rủi ro so với huấn luyện ngoài thực địa.

Trong phạm vi nghiên cứu hiện tại, mô hình động học và các kịch bản thử nghiệm được xây dựng ở mức phù hợp với mục tiêu huấn luyện. Trong tương lai, hướng nghiên cứu có thể tập trung vào mở rộng mô hình cho huấn luyện đa tác nhân, tích hợp các tiêu chí đánh giá kỹ năng học viên và nâng cao độ chính xác của mô hình mô phỏng, nhằm đáp ứng tốt hơn yêu cầu huấn luyện trong thực tế.

TÀI LIỆU THAM KHẢO

- [1] J Hogue, J., Allen, R., MacDonald, J., Schmucker, C., Markham, S., & Harmsen, A. "Virtual reality parachute simulation for training and mission rehearsal," *In 16th AIAA Aerodynamic Decelerator Systems Technology Conference and Seminar*, 2001.
- [2] BiSIM, "PARASIM - Virtual Reality Parachute Training Simulator," 2025.
- [3] Quantum3D, "VPS - Virtual Reality Parachute Simulator," 2025.
- [4] Havelsan, "Havelsan - Parachute Simulator," 2025.
- [5] N. T. Kiên, L. Anh, N. T. Lan, "Nghiên cứu xây dựng hệ thống mô phỏng huấn luyện nhảy dù ứng dụng công nghệ thực tế ảo," in *Hội thảo quốc gia lần thứ XXIV: Một số vấn đề chọn lọc của công nghệ thông tin và truyền thông*, 2021.
- [6] S. A. Forces, "Fact Sheet - Parachute Flight Simulator," Ministry of Defence Singapore, 2018.
- [7] SimCentric, "VR simulator platform trialled by British Armed Forces (including Parachute Regiment)," *Army Technology / UK MoD*, 2020.
- [8] R. S. Sutton, "Learning to Act Using Real-Time Dynamic Programming," *Journal of Artificial Intelligence*, vol. 4, p. 105-131, 1993.
- [9] Kiran, Bangalore & Sobh, Ibrahim & Talpaert, Victor & Mannion, Patrick & Sallab, Ahmad & Yogamani, Senthil & Perez, Patrick, "Deep Reinforcement Learning for Autonomous Driving: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-18, 2021.
- [10] Z. Guo, "A review of decision-making frameworks for autonomous vehicles," *Applied and Computational Engineering*, pp. 142-150, 2024.
- [11] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, Daan Wierstra, "Continuous control with deep reinforcement learning," *arXiv:1509.02971*, 2019.
- [12] Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov "Proximal Policy Optimization Algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [13] A. Z. P. A. S. L. Tuomas Haarnoja, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," *arXiv:1801.01290v2*, 2018.
- [14] H. v. H. D. M. Scott Fujimoto, "Addressing Function Approximation Error in Actor-Critic Methods," *arXiv:1802.09477v3*, 2018.

Creating autonomous control objects for parachute training applications

Nguyen Trung Kien

ABSTRACT

Parachuting is an extreme sport that also serves as an important training component in the military, requiring participants to possess strong psychological composure and master specialized techniques. In practice, parachuting is typically organized as group jumps, with multiple people jumping in the same wave. Therefore, parachutists need to train their coordination abilities and situational handling skills, such as avoiding mid-air collisions, maintaining safe distances, or managing situations when multiple parachutes land in the same area. This paper proposes a model applying artificial intelligence to develop virtual objects that participate in the parachuting process and are capable of automatically controlling the parachute after exiting the aircraft to approach a designated position on the ground. This solution can be integrated into parachute training applications to support training parachutists in practicing group jump missions. The research results demonstrate that integrating artificial intelligence into parachute training applications is feasible and has practical

application potential, contributing to improved training effectiveness and reduced risks compared to traditional training methods.

Keywords: *reinforcement learning, PPO, training simulation, parachuting, AI agent*

Received: 08/01/2026

Revised: 05/02/2026

Accepted for publication: 09/02/2026